

## ETHICAL JUDGMENT AND MOTIVATION

David Faraci (Georgetown) and Tristram McPherson (Ohio State)  
[faraci@gmail.com](mailto:faraci@gmail.com); [dr.tristram@gmail.com](mailto:dr.tristram@gmail.com)

This is the authors' ms. of June 2016, of a paper to be published in the Routledge Handbook of Metaethics. When it becomes available, please refer to and cite the published version of this paper, if possible.

If you face a difficult choice, it is natural to ask yourself *what ought I to do?* And if you answer that question to your own satisfaction, we typically expect you to then perform the action that you've judged you ought to do. Related lines of thought have led many philosophers to conclude that *ethical* judgments – such as the judgment I ought to do such-and-such – have a distinctive connection to motivation and the explanation of action. And many have further taken this connection to have important implications for the nature of ethical judgment.

The literature on this topic has largely focused on the relationship between motivation and *moral* judgment in particular. This is understandable. First, metaethicists have only recently begun distinguishing questions about ethics from questions about morality, and focusing on the former. Second, many endorse claims like:

**Moral Rationalism** Necessarily, if someone *morally* ought to perform an action, she also *ethically* ought to perform it.

Indeed, some claim that this is a conceptual truth (e.g., Smith 1994: 185). This might seem to suggest that ethical and moral judgments bear the same connection to motivation, so that it does not matter which we focus on. However, both Moral Rationalism and this apparent implication are controversial (for discussion, see e.g., Brink 1986; Smith 1994: §3.3; van Roojen 2010; Bromwich 2013; Markovits 2014). In light of this, we think it is fruitful to directly address the relationship between motivation and judgments about what one ethically ought to do. This is both an intrinsically interesting metaethical issue, and a potentially crucial moving part in the dialectic concerning Moral Rationalism and the nature of moral judgments. This

chapter thus focuses primarily on the relationship between ethical judgment and motivation, although we highlight connections to debates over moral judgment and Moral Rationalism where appropriate. It is worth noting that in what follows we often reference papers that focus on moral judgment, when we are discussing an issue concerning ethical judgment. We only flag this difference when we think that this contrast makes a structurally important difference to the point or argument being made in that paper.

We begin by introducing the core motivations for thinking that there is some necessary connection between ethical judgment and motivation, and exploring the consequences of a strong account of the connection, which we call *Simple Internalism* (§1). We then explore the evidence for and against this simple view (§2). Next, we explore three influential ways of modifying the simple view (§3). We examine the case for denying the existence of a necessary connection between ethical judgment and motivation (§4). Finally, we briefly consider several issues about the relationship between mind and language that complicate ordinary ways of discussing this connection (§5).

## 1. An introduction to ethical judgment and motivation

This section does four things. First, it more carefully introduces the initial case for thinking that ethical judgment is linked to motivation. Second, it introduces the strongest form of this connection that has been influential in the literature. Third, it explains how this account of the connection can be used in an argument for a distinctive, *non-cognitivist* account of the nature of ethical judgment. Finally, it introduces and explains some of the central assumptions about motivation and the explanation of action that are required for this argument to work.

To begin, it will be useful to introduce some common (but controversial) assumptions about human psychology that frame our discussion. The first assumption is that much of our thought can be modeled in terms of *attitudes* and *contents*. For example, one might change from *wondering whether one should be vegetarian* to *judging that one should be vegetarian*. This is a change in the attitude (from wondering to judging) one bears towards a single content: that one should be a vegetarian. People

can have many different sorts of attitudes towards a given content: *believing, hoping, fearing, desiring, supposing for the sake of argument*, etc.

Notice that some types of attitude towards an ethical content bear no apparent tie to motivation. For example, there is no reason to expect me to have any particular motivation if I merely wonder whether I ought to perform a certain action. This chapter focuses on a specific attitude-type: ethical *judgment*. Judging is an attitude that is intended to clearly and intuitively contrast with, for example, hoping, fearing, or wondering. As we'll see, the nature of such ethical judgment – whether it is a species of belief, or of desire, or of something else entirely – is one of the central controversies impacted by the debate over the relationship between ethical judgment and motivation.

The contents of attitudes also take many forms. For example, one can have attitudes towards individuals (loving Achmed), properties (seeking happiness), or propositions (worrying that Alice is cruel). Another important category is *self-ascribing* or *de se* content. For example, if Jane thinks Jane is tired, this is an ordinary propositional content that happens to be about her. In the right circumstances, Jane might have this thought without noticing that it is about her (compare Perry 1979). By contrast, if Jane has the thought I am tired, this *de se* thought transparently ascribes tiredness to *the thinker of the thought*.

This chapter focuses on the motivational significance of judgments with self-ascribing ethical content, of the form I ought to perform action A. We assume that such judgments purport to *settle what to do* in a way that contrasts with the deliverances of other normative standards, like etiquette or the rules of a game (for discussion, see the chapter “The Varieties of Normativity”). We can understand Moral Rationalism as the claim that the moral ought has the same authoritative purport. However, we make ethical ought judgments about cases where morality is – at least intuitively – silent. For example, imagine concluding that you will feel more confident today if you wear your favorite shirt. You might conclude that you thus ought to wear it, without thinking that morality bears on this question at all. (For discussion, see the chapter “Ethics and Morality.”)

In focusing on ethical ought judgments in this chapter, we ignore the relationship between motivation and other sorts of ethical judgments. These include

other *strong deontic* judgments, such as that corporal punishment is right or wrong; *evaluative judgments* such as that pleasure is good; *aretaic judgments* such as that courage is virtuous; *thick ethical judgments* such as that charitable giving is generous; and *weak deontic judgements* such as that charitable giving is supererogatory. We also set aside questions about the motivational significance of ethical judgments about *others'* behavior.

It is uncontroversial that moral judgments can take each of the above forms. Ethical judgments can take at least some of these forms, which a full account of the relationship between ethical judgment and motivation will therefore need to address. Such a full theory might significantly complicate the lessons that metaethicists have been inclined to draw on the basis of discussion of self-ascribing ought-judgments (for one example, see Archer *forthcoming*). (For brevity, we call self-ascribing ought-judgments 'ethical judgments', unless a contrast with other ethical judgments is important to the relevant discussion.)

With these clarifications in hand, we can ask: why think that ethical judgment is closely connected to motivation? One answer is suggested by an influential example from Michael Smith:

Suppose we are sitting together one Sunday afternoon. World Vision is out collecting money for famine relief, so we are waiting to hear a knock on the door. I am wondering whether I should give to this particular appeal. We debate the pros and cons of contributing and, let's suppose, after some discussion, you convince me that I should contribute. (Smith 1994: 6)

Smith makes two claims about this case. The first is that, if the canvasser then knocks on the door, you would expect Smith to give. The second is that only certain *special explanations* suffice to dispel this expectation. For example, one excellent explanation would be that Smith has subsequently changed his mind. Another would be that he succumbed to weakness of will.

If correct, these claims make ethical judgments unusual. If all you knew about Smith was that he believed I can contribute to World Vision's mission by giving the canvasser money, it would not be puzzling if Smith then declined to give. We would be equally unsurprised if Smith were unmotivated by certain non-ethical normative judgments, such as a judgment that contributing to World Vision would meet the standards of etiquette or the rules of some game.

These contrasts seem to suggest that ethical judgment is distinctively connected to motivation. One strong thesis is that there is a *necessary* connection between ethical judgment and motivation. Because this connection is often taken to flow from the nature of ethical judgment (and hence to be ‘internal’ to it), we call this view *ethics/motives judgment internalism* (or, for brevity, *internalism*). By parity, we call the contrary view – that there is *no* necessary connection between ethical judgment and motivation – *externalism*. (Compare Darwall 1996: 308 for the classic characterization of judgment internalism.) The internalist claim that there is *some* necessary connection here is importantly imprecise. The next few sections address various precisifications. Begin with:

**Simple Internalism**      Necessarily, if you judge that you ethically ought to do A, you are motivated (to some extent) to do A.

Simple Internalism is poised to help explain both of Smith’s claims. First, if the truth of Simple Internalism were widely implicitly known, that would explain our strong expectation that Smith would give in his scenario. Second, because Simple Internalism merely entails that the ought-judge has *some* relevant motivation, it is compatible with explanations of failure to act that appeal to weakness of will. By contrast, consider the much stronger thesis that if you genuinely judge that you ought to do A, and you are ‘of one mind’ about this, you *will* make the attempt to do A (Gibbard 2003: 153). This stronger thesis can seem implausible exactly in virtue of ruling out weakness of will.

Simple Internalism has been taken to have substantial consequences for the nature of ethical judgments, in light of its relationship to the *Humean theory of motivation*. The core idea of the Humean theory is that for an agent to be motivated to do A, she must desire to do B, and believe that by doing A she will do B. That is, every intentional action is explained by the agent’s belief about what she is doing, and a desire she takes that action – so understood – to fulfill.

Philosophers are often attracted to the Humean theory because they accept a *functionalist* account of the nature of belief and desire. This idea can be illustrated by a famous analogy due to Elizabeth Anscombe (1957: 56). Imagine a detective hired to follow a shopper around a grocery store and find out what he purchases. Suppose that both the detective and the shopper have lists of groceries. The shopper’s list functions to *guide* his behavior: when all goes well, if ‘butter’ is on the list, he puts butter in his

cart. The shopper's aim is to make the contents of his cart conform to the contents of his list. The detective's list, by contrast, functions to *represent* the world: when all goes well, if there is butter in the cart, the detective adds 'butter' to her list. Her aim is to make her list conform to the contents of the shopper's cart.

Psychological states can be understood in similar functional terms. A *cognitive* or *belief-like* psychological state functions like the detective's list: when all goes well, one will *believe* that there is butter in the shopper's cart only when there is. A *non-cognitive* or *desire-like* psychological state functions like the shopper's list: when all goes well, the desire to put butter in the cart will motivate one to put butter in the cart. As the Humean theory suggests, part of *all's going well* here is the presence of a relevant means-end belief: for example that picking up the butter and dropping it thusly will get it into the cart. If the shopper instead believes that the way to get butter into his cart is to throw it at the wall, he will not tend to wind up with butter in his cart. Crucially, the functionalist Humean insists that the cognitive and non-cognitive functions are *distinct*: for example, while the desire that there be butter in the cart motivates you to put butter in the cart, it would be bizarre for the belief that there is butter in the cart to motivate you to put butter in the cart. (For two approaches to defending the Humean theory of motivation, see Smith 1987 and Sinhababu 2009. For challenges to the Humean Theory or its typical functionalist gloss, see, e.g., Little 1997; Scanlon 1998; Sobel and Copp 2001; Coleman 2008; Schueler 2009; Frost 2014.)

It is important to emphasize that the issue here is about functionally characterized states. Despite their names, it is controversial whether these functional states map neatly onto the attitudes we typically mean by 'belief' and 'desire'. (For some complications, and different ways of understanding the distinction between cognitive and non-cognitive states, see the chapter "Cognitivism and Non-Cognitivism.")

Combined with Simple Internalism, the Humean theory underwrites an elegant argument for *non-cognitivism* about ethical judgment: the thesis that ethical judgments are *desire-like* attitudes. If only desire-like attitudes can motivate, and ethical judgments always motivate, then wherever there is an ethical judgment, there must be a non-cognitive attitude. Arguably, the simplest explanation for this partnership is that ethical judgments *are* non-cognitive attitudes. Of course, this isn't the *only* possible

explanation. For example, a cognitivist could propose to explain Simple Internalism by citing a universal desire to do what one ought to do. But why think that, *necessarily*, every ethical judge has such a desire? In light of questions like this, many philosophers have thought that non-cognitivism is the *best* explanation for theses like Simple Internalism.

## 2. The evidence for and against Simple Internalism

This section considers evidence for and against Simple Internalism, as well as resources that the simple internalist can use to rebut evidence against the view. It is important to emphasize in this context that Simple Internalism is in one respect an extremely *strong* view, because it claims that motivation to do A *necessarily* accompanies the judgment that one ought to do A. Because of this, showing that *in many cases* motivation accompanies such ethical judgments is not sufficient to establish the view; something more must be done to show that this connection is necessary.

A range of evidence for Simple Internalism is suggested by reflection on possible cases. Suppose, for example, that someone's avowed ethical claims fail to line up with his behavior. Which of these – the claims or the behavior – would you consider the better guide to his ethical judgments? Arguably his *behavior* is the better guide (compare Hare 1952: 1). But the idea that behavior is a good guide to ethical judgment only makes sense if ethical judgment tends to be connected to motivation. And this tendency must be quite strong, in order for the evidence of this connection to outweigh the ordinary presumption of sincere utterance. Smith's claims about the famine relief example, discussed in the previous section, provide complementary evidence: if (as Smith claims) we strongly expect motivation to accompany a *new* ethical judgment (like the one Smith makes in the example), this supports the idea that we take motivation to be connected to ethical judgment itself. (For empirical evidence that the sorts of expectations Hare and Smith suggest are widely shared, see Eggers 2015.)

One limitation of the sort of evidence offered by Hare and Smith is that it is *indirect*: if successful, it most directly identifies our widely shared tacit *beliefs* about ethical judgments, rather than providing direct evidence about the nature of the

judgments themselves (for a discussion that exploits this distinction, see Braddon-Mitchell 2006). One might thus attempt to complement this work by studying the nature of ethical judgments directly. Consider one discussion of *moral* judgment that provides an instructive model. Jesse Prinz (2015: §4.2.1) points to three converging lines of evidence for internalism about moral judgments: (1) that moral cognition is correlated with heightened emotion, (2) that elicited emotions can alter moral judgments, and (3) that emotional differences and deficits are correlated with differences and deficits in moral judgment. Prinz suggests that this evidence is *best explained* by a sentimentalist view on which moral judgments consist of emotional attitudes.

The form of Prinz's argument answers a central challenge to the possibility of empirical arguments for Simple Internalism: Simple Internalism is a strong modal claim, and it may be hard to see how empirical evidence alone could support such a claim. Prinz's account suggests an answer: he argues that the empirical evidence supports sentimentalism, which is a constitutive claim about the nature of moral judgments. And constitutive accounts characteristically have strong modal consequences.

What can be said *against* Simple Internalism? As we noted above, this view posits a *necessary* connection. Its modal strength thus makes it a natural target for counterexamples: establishing a single possible case of ethical judgment without motivation would suffice to refute it. One prominent style of counterexample concerns persons suffering from a quite general lack of motivation, such as deeply depressed persons. It might seem completely unsurprising that a depressed person could sincerely make an ethical judgment and yet fail wholly to be motivated by it (Stocker 1979: 744).

One line of reply to this style of objection takes advantage of a respect in which Simple Internalism is a *weak* claim. It requires only that one possess *some* motivation to do what one judges that one ought to (Finlay 2004: 209). And this makes it possible to resist putative counterexamples like Stocker's depressed person, by suggesting that the motivation exists, but is *weak*, perhaps to the extent of being unnoticeable even to the agent himself.



Another prominent style of counterexample to Simple Internalism appeals to the possibility persons who simply do not care about ethics. Consider Plato's Thrasymachus, who believes (roughly) that justice is what is in the interests of the powerful, and is therefore not moved at all by the thought of what justice demands of him. Notice that this sort of view is easiest to make plausible with relatively 'thick' moral concepts, like JUSTICE. It is a bit harder with MORALLY OUGHT, although here imagining someone who rejects Moral Rationalism – and hence believes that he *ethically* ought not to act *morally* – can make the case more plausible (compare Brink 1986).

Things get harder still with ETHICALLY OUGHT (compare Ridge 2014: 55-6). There has not been a great deal of effort in the literature to spell out an analogue for Thrasymachus with respect to ETHICALLY OUGHT (for one exception see Greenberg 2009: 156-8). But there are interesting resources to be explored. For example, someone might have an *alienating* conception of what she ethically ought to do: a conception on which what she ought to do simply runs roughshod over everything she most cares about. Such a person might conclude that doing what one ought to do is simply *awful*, and it might not be surprising if such a person had no motivation to do what she judged she ought to. Call such a character *anethical*.

One influential way of replying to this style of counterexample begins by noting that not all sincere claims that deploy ethical *words* thereby express the speaker's ethical *judgments*. For example, we sometimes use ethical words to talk about conventional ethical views, or some specific salient ethical view other than our own. R. M. Hare (1952: 124) noted that in print, we can use *inverted commas* around a word to signal such a use. For example, someone might convey her distaste for local mores by writing: "Around here, 'justice' involves a *remarkable* number of public beatings." Interpreted in this way, Thrasymachus might be understood as claiming that 'justice' – i.e. *what other people call 'justice'* – amounts to the interests of the powerful. Crucially, when Thrasymachus makes these claims, he does not thereby directly express his own justice-judgments. His lack of motivation when making such claims would therefore be entirely compatible with Simple Internalism. (For arguments against the plausibility of this strategy as a reply to the relevant range of cases, see Svavarsdóttir 1999: 188-192; for a different internalist reply to this style of case, see Bromwich 2013.)

Psychopaths have received considerable attention in recent discussion of these issues, as potential real-world examples of Thrasymachus-style *amoralists*. A significant part of the literature on psychopaths and internalism concerns whether psychopaths should be interpreted as making genuine moral judgments, or something more like Hare-style inverted commas judgments (e.g., Roskies 2003; Nichols 2004; Prinz 2007; Kennett and Fine 2008; Kumar 2016a).

The significance of this literature for Simple Internalism, as a view about *ethical* judgment, is less clear. The first difficulty is that psychopaths seem indifferent to morality, not to their ethical judgments per se, and there are few other candidates for real-life anethical judges. Of course, those who endorse Moral Rationalism will maintain that amoral judges are *ipso facto* anethical judges. But if psychopaths make genuine moral judgments, this might more naturally be taken to cast doubt on Moral Rationalism itself, as opposed to threatening Simple Internalism.

Ironically, a final way to challenge Simple Internalism can be developed from the fact that motivational states are functional states. Functional states are usually analyzed in terms of dispositions, and it is possible for a disposition to be present but *masked*. For example, a vase can have the dispositional property of being *fragile*, where this property is *masked* by its being packed safely in a box (Johnston 1992). One might think that a condition like depression could mask the motivational force of some of one's ethical judgments. If so, then even if ethical judgments are desire-like functional states, Simple Internalism would be false.

A substantial number of philosophers have thought that – in part in virtue of the sorts of cases explored in this section – ethical judgment is possible without motivation. However, many of these philosophers have also thought that the simple internalist is correct in thinking that there is *some* sort of necessary connection between ethical judgment and motivation. We now turn to explore such views.

### 3. Defeasible internalisms

In this section, we introduce three important ways of developing the idea that the connection between ethical judgment and motivation is necessary but not universal.

These accounts take the relationship to be mediated by *rationality*, *normalcy*, and *linguistic community*, respectively.

### 3.1 Rationality internalism

The first defeasible form of internalism we discuss is:

**Rationality Internalism** Necessarily, if you judge that you ethically ought to do A, and you are structurally rational, you are motivated (to some extent) to do A.

This formulation amends Simple Internalism by adding an additional condition: structural rationality. The inclusion of ‘structurally’ here is important. To see why, consider the claim that in order to count as *substantively* rational you must be motivated to do what you judge that you ought to. This might be interpreted as telling us a lot about what is required to count as substantively rational, but little about the nature of ethical judgment (compare Schroeter 2005). This worry doesn’t seem to apply to Rationality Internalism, because we have an independent grasp on structural rationality as a matter of the joint coherence of one’s attitudes. For example, if you believe both P and not-P, then you are structurally irrational. Rationality Internalism, then, differs from Simple Internalism in proposing that ethical judgment without motivation is *possible*, but only if the agent in question is in some way incoherent.

One motivation for Rationality Internalism begins by noting that the judgment I ought to do A has its natural home in deliberation, arguably the paradigmatic mechanism for rational governance of one’s own actions. This might make it seem plausible that the failure of motivation following an ought-judgment is structurally irrational, in the same way that it might seem plausible that a failure to apportion one’s beliefs to what one takes to be the weight of one’s evidence seems to be structurally irrational (for relevant discussion, see Pettit and Smith 1996; Burge 2000; and Schroeter 2005.) Notice that this motivation is clearer than in the case of moral judgment, where the (im)plausibility of Moral Rationalism is another crucial moving part.

Rationality Internalism has resources to address some of the other alleged counterexamples to Simple Internalism, discussed in the previous section. Consider the depressed person. The lack of motivation imagined here is quite *general*: this person will also struggle to do what he judges that he *would like to do*, or what he judges *would*

*make him feel better*. This seems like a good candidate for a rational failing: this person's motivational system fails to cohere with his own perspective on the world. Or consider the psychopath. It has been argued that psychopaths have significant deficits in practical reasoning (Nichols 2004 and especially Maibom 2005). This means that even if psychopaths *do* make genuine moral judgments, their existence may be compatible with both Rationality Internalism *and* Moral Rationalism, if their lack of motivation can be connected to their structural irrationality. Rationality Internalism is also poised to explain the sorts of claims that motivated Simple Internalism: the sorts of expectations that Smith and Hare pointed to can be understood as reasonable, given a background assumption that we are ordinarily (approximately) structurally rational.

What consequences does Rationality Internalism have for our theories of ethical judgment? The first thing to notice is the contrast with the non-cognitivist's explanation of Simple Internalism. According to Rationality Internalism, it is possible to have an ethical judgment without motivation. This rules out identifying ethical judgment with any necessarily motivating state. If every token non-cognitive state were necessarily motivating, this would spell trouble for the non-cognitivist. However, we should not think this (Björnsson 2002: §4). For example, Ridge (2015: 145-6) argues that *general plans* are an example of non-cognitive states that are generally but not necessarily connected to motivation. So Rationality Internalism does not *rule out* non-cognitivism.

On the other hand, the straightforward case from Simple Internalism *against cognitivism* fails to carry over. Recall that case: on Simple Internalism, ethical judgment guarantees the presence of motivation, and according to the Humean theory of motivation, no cognitive state by itself can guarantee motivation. Given these assumptions, the best explanation of ethical motivation seemed to rule out cognitivism about ethical judgment. But Rationality Internalism rejects the guarantee of motivation, so it cannot support a parallel argument.

Instead, if Rationality Internalism has implications for this debate, they turn on the question: which kinds of states can succeed or fail to cohere with motivational states? Both cognitivists and non-cognitivists can appeal to the plausibility of so-called *enkratic* principles of structural rationality, such as: *if you judge that you ought to do A, then*

*you intend to do A* (Broome 2013). Can one side of the cognitivism/non-cognitivism debate offer an explanation for the truth of such a principle that is unavailable to her opponent? Consider two attempts.

A simple cognitivist theory might analyze the judgment I ought to do A as a belief with the content: *were I structurally rational, I would do A*. This sort of analysis seems poised to explain why the enkaptic principle is a principle of structural rationality, in virtue of its content. (For a sophisticated proposal along these lines, see Smith 1994: Ch. 5.) If an analogous account is unavailable to the non-cognitivist (something we take no position on here), this would support a Rationality Internalism-based argument for cognitivism.

Conversely, it might be thought that the non-cognitivist has the advantage here, since on her view ethical judgments are intrinsically apt to motivate, even if they do not do so in every case. But it is not obvious how this grounds rational links. For example, one might think that the ability to *mask* the motivational force of one's non-cognitive states is sometimes rational: consider someone who is able to experience anger, but control what, if anything, that anger disposes him to do.

### 3.2 Normality internalism

A different way to retain the core internalist idea, while granting the force of the apparent counterexamples (such as the anethical judge and the depressed person) is to focus on the idea that in these examples, the unmotivated judges are *abnormal*. We could state the proposal this way:

**Normality Internalism** Necessarily, if you judge that you ethically ought to do A, and you are normal, you are motivated (to some extent) to do A.

It is crucial that the notion of normality at play here is not intended as statistical. Rather, the key idea is that we can only imagine cases of unmotivated ethical judgment as parasitic on a robust pattern of connection between ethical judgment and motivation (compare Dreier 1990). For example, the depressed person is imagined to have acquired ethical concepts, been motivated in the ordinary way, and only then fallen into a depression that deprives him of motivation. And Thrasymachus is imagined to have started out motivated to do what is just, and transitioned to his unmotivated state as a result of his investigation into the nature of justice.

One way to motivate Normality Internalism is to imagine someone who is raised without ethical concepts, then joins a society that possesses ethical concepts. He learns how to follow along in conversations, how people use ‘ought’, etc., but he never experiences any motivation to do what he says he judges he ought to. It may be more tempting to ascribe an inverted commas meaning – as opposed to a genuine ethical judgment – to this person’s use of ‘ought’ than to do so with the depressed person. The normality condition can explain why: this person – unlike the depressed person – has never instantiated the normal connection between ethical judgment and motivation.

Normality Internalism can explain some of the claims that motivated Simple Internalism, but only given background empirical assumptions. For example, because normality is *not* a statistical notion, the expectations that Smith and Hare mention will only be vindicated when we believe that the relevant speakers are normal. The normality internalist might insist that this is as it should be: when we learn any of a variety of things about how an ethical judge has become abnormal, the Smith or Hare-style expectations of that judge tend to disappear. Normality Internalism also arguably has more resources to address counterexamples than Rationality Internalism. For example, the normality internalist does not need to diagnose some structural irrationality in the anethical judge, in order to defend her view.

As these points bring out, the Normality view is a recognizably more modest version of internalism. One might worry that this modesty is purchased at the price of opacity: while we have an initial gloss and motivating examples to orient us to the relevant notion of normality, it is not clear how precisely to understand that notion, or how explanatorily interesting it is (compare Svavarsdóttir 1999: 175, n.7).

It is an interesting question what the upshot of this view is for the debate between cognitivists and non-cognitivists. As with Rationality Internalism, there is no simple argument for non-cognitivism (for relevant discussion see Strandberg 2012 and Toppinen 2015). For example, if all abnormal conditions could be understood as conditions that mask an underlying non-cognitive disposition, this would be grist for the non-cognitivist’s mill. But it is unclear why we should think that all abnormal ethical judges retain a motivational disposition. On the other hand, Normality Internalism leaves the cognitivist needing to explain what sort of cognitive state could

be such that in order to *get into it*, one must have a certain pattern of motivations, even if one can then go on to lose that pattern. One possibility is that ethical judgment is a *hybrid* state, consisting of cognitive and non-cognitive elements that are ordinarily linked by robust causal mechanisms (e.g., Kumar 2016b). Cases of abnormality might be explained as cases in which these sorts of normal mechanisms break down, or are masked by other psychological processes (for a different cognitivist explanation, see Dreier 1990).

### 3.3 Community-level internalism

The final way of refining internalism that we discuss sets aside the individual and focuses instead on the idea that there is a necessary connection between ethical judgment and motivation at the level of the linguistic community. Suppose you find yourself confronted with an alien linguistic community, *Anethicalia*, whose members happen, through the infinite wackiness of the cosmos, to speak a language that is identical, phonetically and grammatically to English, and also semantically *very* similar. Their planet has tall plants they call ‘trees’; small, furry animals they call ‘cats’ and ‘dogs’; emotions between partners they call ‘love’; and so on. They often talk of what they ‘ought’ to do, much as we do. They often claim that they ‘ought’ to save the dying, avoid hurting others, give to World Vision, and care for their children. There is just one striking contrast between us and the Anethicalians: they are utterly unmotivated by these judgments.

Some philosophers have suggested that linguistic communities like this are *not* deploying ethical concepts (e.g., Lenman 1999 and Bedke 2009; for a contrary view see Gert and Mele 2005). The idea of community-level internalism, then, is that, necessarily, there is a *community-level* connection between ethical judgment and motivation, for any community of speakers that uses words to conventionally voice their ethical judgments. Stronger and weaker versions of this sort of view are possible. A stronger version would insist that necessarily, in any linguistic community which has a term in its language that means *ethically ought*, speakers in that community are normally motivated to do A when they judge that they ethically ought to do A. By contrast, a weaker version of the view would insist that a community of normal ethical judges might retain their ethical concepts, even as they evolved into anethical judges.

The internalist connection would be retained by insisting that there could not be a community that deployed ethical concepts that had *always* been overwhelmingly anethical (Bedke 2009; cf. Dreier 1990).

Like the Normality view, the community-level account has the virtue of smoothly accommodating many sorts of unmotivated ethical judges, provided such judges do not constitute a whole community. However, because the community-level view posits a relatively tenuous link between ethical judgment and motivation, it may have more difficulty vindicating the stronger sorts of claims that Smith and Hare used to motivate internalism in the first place. In light of this, the view is perhaps best understood as motivated by distinct considerations, like the cases introduced in this section.

Critics can attempt to debunk the cases that are supposed to support the community-level view. Consider one such approach, which begins by noting that on a plausible substantive account of ethical facts, such facts will be a function of something like prudence, cooperation, special relationships, satisfaction of our goals, etc. (or a subset of these things). Now consider a simple form of externalist cognitivism, on which ethical judgment consists in beliefs *about* prudence, cooperation, etc., with no necessary connection to motivation. (Focusing on this view serves only to make the point vivid; similar points hold for other views.) On this view, why would members of a community tend to make ethical judgments, or, more broadly, deploy ethical concepts at all? Presumably because they *care* about prudence, cooperation, etc. enough to want to keep track of them. This drives home the point that the case proposed to support community-level internalism would have to be quite bizarre: *First* we would need to imagine a functioning society where the members did not typically care about prudence, cooperation, etc. This is already difficult. Then we would need to imagine that such a society nonetheless got a bit of discourse up and running to talk about these topics. This is even more puzzling. Then we are supposed to be confident that this society does not make ethical judgments. Even if we have these reactions about such bizarre cases, it is not wholly clear whether we should take them as any kind of evidence (for worries about the evidential significance of related cases, see Dowell 2015).



The relationship between community-level internalism and theories of the nature of ethical judgment is even less clear than on the Normality view. Both cognitivism and non-cognitivism, as we have spelled them out, are claims about the nature of the *individual psychological states* that constitute ethical judgment. But the community-level view suggests that we can tolerate considerable variation in the nature of such states, provided they are related appropriately to a broader communal pattern (we return to this point in §5).

#### 4. Denying a necessary connection

*Ethics/motives judgment externalists* deny a necessary connection between ethical judgment and motivation. Externalists have two central tasks in arguing for their view. First, they must argue against the presence of a necessary connection. This task has already largely been explored in this chapter: many of the challenges to each form of internalism explored above can also be understood as *prima facie* evidence for externalism. Together, the range of such cases might be taken to constitute a powerful argument for externalism.

Whether or not there is a *necessary* connection between ethical judgment and motivation, it is uncontroversial that there is typically a *strong* correlation between these states. The externalist's second task is therefore to explain this correlation without appealing to such a necessary connection. One natural explanation is that many people – especially generally virtuous people – find appealing the idea of acting ethically; they desire *to do what they ought to do*, under that very description. Michael Smith complains that this is not the sort of motivation a good person would have (1994: §3.5). For example, a good person who judges that she ought to help you should be motivated by *the fact that you need help*, not simply by the desire *to do what she ought to*. (For replies to Smith on this point, see e.g. Strandberg 2007, and several of the papers cited therein.)

One straightforward externalist reply to Smith points out that we should expect ordinary good people to have a *plurality* of relevant motives that are contingently but strongly correlated with what they believe they ought to do, and so are apt to explain their acting in accordance with such ethical judgments. For example, many people are motivated by prudence or kindness or respect, or by direct concern

for certain individuals in their lives. Many people also want to act justifiably, and believe that ethical action is a way of doing so.

Given this, the externalist can appeal to an *overlapping set* of these sorts of commonplace motivations to explain the robust correlation between ethical judgment and motivation. This might seem ideally suited to explain the phenomenon in question. On the one hand, by appealing to such an overlapping patchwork of motives that can be expected to vary between people, this picture can explain the wide variation in the nature, strength, and resilience of ethical motivation between people (noted by Svavarsdóttir 1999: 161). On the other, we would expect ordinary good people to possess most of the motives just mentioned: this is part of what makes them good!

This picture can also be used to explain why only certain special explanations will *ordinarily* make lack of ethical motivation intelligible. For in many ordinary cases, what someone judges that she ought to do is correlated with a wide variety of substantial emotions. But this is not always the case. Consider a fact that is familiar to every vegan philosopher. We frequently meet people who say: “I am convinced that I ought to be vegan, but I am just not motivated to do it.” Yet we often take them to be sincere in these judgments. Arguably, this is because we recognize that such people are normally motivated by many of their *other* ethical judgments, but we also understand how difficult it can be to emotionally connect with distant animal suffering or to *feel* wrongness in apparently mundane activities like drinking milk. The patchwork picture of ethical motivation suggested in the previous paragraph is well-suited to vindicate these natural thoughts.

## **5. Complications: mind and language**

Much of the discussion of ethical judgment and motivation takes as evidence claims about the relationship between what people *say*, and what they are motivated to do. But notice that facts about what people say are first and foremost facts about language, not psychology. In this section, we note three ways in which focusing on language as a crucial moving part can complicate the discussion offered thus far.

First, consider whether your favored connection between ethical judgment and motivation (if any) really demands *any* psychological explanation in terms of the nature of ethical judgment. Jon Tresan (2006) has shown that even strong internalist connections do not obviously demand this. To see this, consider the state *Tuesday belief*. ‘Tuesday belief’ just means: any belief held on a Tuesday. There are surely Tuesday beliefs, and necessarily, if one has a Tuesday belief, one has it on a Tuesday. But it would be silly to think that we need to give a *psychological* explanation of why this mental state is necessarily instantiated on only one day of the week. The simple internalist claim that ethical judgment is necessarily accompanied by motivation (e.g.) might get the same treatment: it might be suggested that a certain ordinary belief only counts as an ethical judgment when accompanied by motivation. If this suggestion were right, then just as in the Tuesday belief case, there would be no special psychological puzzle about ethical judgments; they could be ordinary beliefs that cease to count as ethical judgments when their contingent connection to motivation happens to fail.

Second, consider a plausible explanation for the fact that moral and, more broadly, ethical discourse is a cultural universal: such discourse is a means of achieving coordination, shared reasoning, and influence. A functional rationale for treating someone as an ethical judge, then, is that you think of their relevant judgments as apt candidates for such discursive coordination, shared reasoning, and influence. This, in turn, requires that you take yourself to have some reasonable chance of influencing your interlocutors’ behavior through ethical engagement. If correct, this picture might explain many of our judgments about who is making genuine ‘ethical judgments’. But there may be no *specific* psychological states that our interlocutors would need to be in to make such engagement possible (compare Björnsson and McPherson 2014).

Third, consider a distinctive commitment of the community-level views, suggested at the close of §3.3. These views seem committed to a sort of language-first ‘anti-individualism’ concerning ethical judgments (cf. Burge 1979). Community-level internalists appear committed to the idea that two individuals could have qualitatively identical brain states and dispositions, and yet the first could count as making ethical judgments while the second does not, because the second individual is not located in a linguistic community with the relevant properties. This consequence is inconsistent

with common assumptions about how to individuate ethical judgments. For example, we sketched the distinction between cognitivism and non-cognitivism in terms of the functional natures of elements of individual psychologies. The anti-individualist picture suggests that it may be a category mistake to individuate ethical judgments in this way. If community-level evidence is compelling, it may thus warrant a shift of focus from *psychological* questions about the nature of ethical judgment to *metasemantic* questions about how best to explain the meaning of ethical words in a public language.

On the other hand, anti-individualism is a controversial feature of the community-level view, because many philosophers are tempted to think that ethical judgments are distinctively apt for psychology-level – as opposed to community-level – individuation conditions. Suppose that Sally lives in Anethicalia (the community mentioned in §3.3). But suppose that, for whatever reason, her judgments about what she ‘ought’ to do play a central role in her agential life: she often plans for the future by thinking about what she ‘ought’ to do, she is robustly motivated by her judgments about what she ‘ought’ to do, etc. It may seem implausible to insist that she fails to make ethical judgments simply because the *words* she uses to express the relevant judgments are not conventionally associated with any motivational response. This might be taken as evidence that internalism really *is* a thesis about the nature of ethical judgment, rather than about how we use language.

### Conclusions

It is a striking fact that our ethical judgments appear more intimately connected to motivation than many of our other judgments. Many metaethicists have taken this fact to provide the seeds of powerful arguments for distinctive accounts of the nature of ethical judgment. This chapter has aimed to introduce readers to the complexity of the issues facing these arguments, and to some of the tools that philosophers have deployed in seeking to understand that complexity. We hope this introduction is useful to readers interested in further investigating these important issues

### Acknowledgements

Many thanks to Jamie Fritz and David Plunkett for helpful comments on previous drafts.

### References

- Anscombe, G.E.M. 1957. *Intention*. Harvard University Press.
- Archer, A. forthcoming. "Motivational Judgement Internalism and the Problem of Supererogation." *Journal of Philosophical Research*.
- Bedke, M.S. 2009. "Moral Judgment Purposivism: Saving Internalism from Amoralism." *Philosophical Studies* 144 (2): 189–209.
- Björnsson, G. 2002. "How Emotivism Survives Immoralists, Irrationality, and Depression." *The Southern Journal of Philosophy* 40 (3): 327–344.
- Björnsson, G., and T. McPherson. 2014. "Moral Attitudes for Non-Cognitivists: Solving the Specification Problem." *Mind* 123 (489): 1–38.
- Braddon-Mitchell, D. 2006. "Believing Falsely Makes It So." *Mind* 115 (460): 833–866.
- Brink, D.O. 1986. "Externalist Moral Realism." *The Southern Journal of Philosophy* 24 (S1): 23–41.
- Bromwich, D. 2013. "Motivational Internalism and the Challenge of Amoralism." *European Journal of Philosophy*.
- Broome, J. 2013. *Rationality Through Reasoning*. John Wiley & Sons.
- Burge, T. 1979. "Individualism and the Mental." *Midwest Studies in Philosophy* 4 (1): 73–121.
- . 2000. "Reason and the First Person." In *Knowing Our Own Minds*, edited by C. Wright, B. Smith, and C. Macdonald. Oxford University Press.
- Coleman, M.C. 2008. "Directions of Fit and the Humean Theory of Motivation." *Australasian Journal of Philosophy* 86 (1): 127–139.
- Darwall, S. 1996. "Reasons, Motives, and the Demands of Morality: An Introduction." In *Moral Discourse and Practice: Some Philosophical Approaches*, edited by S. Darwall, A. Gibbard, and P. Railton. Oxford University Press.

- Dowell, J. L. 2015. "The Metaethical Insignificance of Moral Twin Earth." In *Oxford Studies in Metaethics*, edited by R. Shafer-Landau. Vol. 11. Oxford University Press.
- Dreier, J. 1990. "Internalism and Speaker Relativism." *Ethics* 101 (1): 6–26.
- Eggers, D. 2015. "Unconditional Motivational Internalism and Hume's Lesson." In *Motivational Internalism*, edited by G. Björnsson, C. Strandberg, R. Francén Olinder, J. Eriksson, and F. Björklund. Oxford University Press.
- Finlay, S. 2004. "The Conversational Practicality of Value Judgement." *The Journal of Ethics* 8 (3): 205–223.
- Frost, K. 2014. "On the Very Idea of Direction of Fit." *Philosophical Review* 123 (4): 429–484.
- Gert, J., and A. Mele. 2005. "Lenman on Externalism and Amoralism: An Interplanetary Exploration." *Philosophia* 32 (1): 275–283.
- Gibbard, A. 2003. *Thinking How to Live*. Harvard University Press.
- Greenberg, M. 2009. "Moral Concepts and Motivation." *Philosophical Perspectives* 23 (1): 137–164.
- Hare, R.M. 1952. *The Language of Morals*. Oxford Clarendon.
- Johnston, M. 1992. "How to Speak of the Colors." *Philosophical Studies* 68 (3): 221–263.
- Kennett, J., and C. Fine. 2008. "Internalism and the Evidence from Psychopaths and 'Acquired Sociopaths.'" In *The Neuroscience of Morality: Emotion, Disease, and Development*, edited by W. Sinnott-Armstrong. Vol. 3. Moral Psychology. MIT Press.
- Kumar, V. 2016a. "Psychopathy and Internalism." *Canadian Journal of Philosophy* 46 (3): 318–45.
- . 2016b. "The Empirical Identity of Moral Judgement." *The Philosophical Quarterly*, 1–22.
- Lenman, J. 1999. "The Externalist and the Amoralist." *Philosophia* 27 (3): 441–457.
- Little, M.O. 1997. "Virtue as Knowledge: Objections from the Philosophy of Mind." *Noûs* 31 (1): 59–79.
- Maibom, H.L. 2005. "Moral Unreason: The Case of Psychopathy." *Mind & Language* 20 (2): 237–257.
- Markovits, J. 2014. *Moral Reason*. Oxford University Press.

- Nichols, S. 2004. *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford University Press.
- Perry, J. 1979. "The Problem of the Essential Indexical." *Noûs* 13 (1): 3–21.
- Pettit, P., and M. Smith. 1996. "Freedom in Belief and Desire." *The Journal of Philosophy* 93 (9): 429–449.
- Prinz, J. 2007. *The Emotional Construction of Morals*. Oxford University Press.
- . 2015. "An Empirical Case for Motivational Internalism." In *Motivational Internalism*, edited by G. Björnsson, C. Strandberg, R. Francén Olinder, J. Eriksson, and F. Björklund. Oxford University Press.
- Ridge, M. 2014. *Impassioned Belief*. Oxford University Press.
- . 2015. "Internalism: Cui Bono?" In *Motivational Internalism*, edited by G. Björnsson, C. Strandberg, R. Francén Olinder, J. Eriksson, and F. Björklund. Oxford University Press.
- Roskies, A. 2003. "Are Ethical Judgments Intrinsically Motivational? Lessons from 'Acquired Sociopathy.'" *Philosophical Psychology* 16 (1): 51–66.
- Scanlon, T.M. 1998. *What We Owe to Each Other*. Belknap Press of Harvard University Press.
- Schroeter, F. 2005. "Normative Concepts and Motivation." *Philosopher's Imprint* 5 (3): 1–23.
- Schueler, G. F. 2009. "The Humean Theory of Motivation Rejected." *Philosophy and Phenomenological Research* 78 (1): 103–122.
- Sinhababu, N. 2009. "The Humean Theory of Motivation Reformulated and Defended." *Philosophical Review* 118 (4): 465–500.
- Smith, M. 1987. "The Humean Theory of Motivation." *Mind* 96 (381): 36–61.
- . 1994. *The Moral Problem*.
- Sobel, D., and D. Copp. 2001. "Against Direction of Fit Accounts of Belief and Desire." *Analysis* 61 (269): 44–53.
- Stocker, M. 1979. "Desiring the Bad: An Essay in Moral Psychology." *The Journal of Philosophy* 76 (12): 738–753.
- Strandberg, C. 2007. "Externalism and the Content of Moral Motivation." *Philosophia* 35 (2): 249–260.

- . 2012. “Expressivism and Dispositional Desires.” *American Philosophical Quarterly* 49 (1): 81–91.
- Svavarsdóttir, S. 1999. “Moral Cognitivism and Motivation.” *The Philosophical Review* 108 (2): 161–219.
- Toppinen, T. 2015. “Pure Expressivism and Motivational Internalism.” In *Motivational Internalism*, edited by G. Björnsson, C. Strandberg, R. Francén Olinder, J. Eriksson, and F. Björklund. Oxford University Press.
- Tresan, J. 2006. “De Dicto Internalist Cognitivism.” *Noûs* 40 (1): 143–165.
- van Roojen, M. 2010. “Moral Rationalism and Rational Amoralism.” *Ethics* 120 (3): 495–525.

### **Suggested further reading**

- Björnsson, G., C. Strandberg, R. Francén Olinder, J. Eriksson, and F. Björklund (eds.). 2015. *Motivational Internalism*. Oxford University Press.
- Rosati, C. S., “Moral Motivation,” in *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2014/entries/moral-motivation/>.